DOCUMENT RESUME

ED 167 188 **IR 007 187**

AUTHOR	Thompson, M. E.
TITLE	Armchair Survey Sampling: An Aid in Teaching Survey Design.
PUB DATE	6 Apr 78
NOTE	11p.; Paper presented at the Symposium on CCMIT (Waterloo, Ontario, Canada, April 6, 1978); For entire proceedings, see IR 007 100
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Audiovisual Aids; *Computer Assisted Instruction; *Educational Strategies; Educational Television; *Higher Education; Individualized Instruction; Instructional Design; *Learning Modules; *Sampling; Simulation; Teaching Methods
IDENTIFIERS	*Computerized Multimedia Instructional Television

ABSTRACT

A fictitious community of 583 households was set up to simulate a survey population, and was used in two laboratory assignments where students "interviewed" householders by a quota sampling procedure and tested the performance of several probability sampling designs. (Author/JEG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167188

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ARMCHAIR SURVEY SAMPLING:
AN AID IN TEACHING SURVEY DESIGN

by

M.E. Thompson

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary E. Thompson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR007187

ARMCHAIR SURVEY SAMPLING: AN AID IN TEACHING SURVEY DESIGN

by

M.E. Thompson
Associate Professor
Department of Statistics

Abstract: A fictitious community of 583 households was set up to simulate a survey population. In two laboratory assignments, students "interviewed" householders by a quota sampling procedure, and tested the performance of several probability sampling designs.

Introduction:

When the author was invited to participate in the COMIT project, the Department of Statistics at the University of Waterloo had just begun to offer a service course called Principles of Survey Design. This course was aimed primarily at senior undergraduates in the social sciences, who are sometimes required to conduct surveys as part of their academic training, and who are frequently faced with the job of interpreting or assessing the results of surveys. The designers of the course were hoping that the fundamental ideas of sampling theory could be communicated to a reasonable proportion of these students via examples and case studies and with a minimum of mathematics. The COMIT project offered an opportunity to enrich and reinforce this program; and it was decided to try to use it to set up "labs" in which students would conduct simulated surveys at the terminals, and in this way gain some practical experience and insight into design alternatives.

Among the basic notions which were selected as being possible and worthwhile to try to illustrate were the following:

(i) the advantages and pitfalls of quota sampling;
(ii) the concept of sampling error in probability sampling;
(iii) the advantages of stratification; and
(iv) the inefficiency of a 'clustered' sampling design compared with a 'dispersed' sampling design.

All these notions were to be discussed at some length in the classroom before the students encountered them in laboratory sessions at the terminals. After the sessions, the students' results were to be discussed further in class. Thus the COMIT module would assume the role of a traditional problem assignment, except that it would make possible the handling of realistic data without requiring a great deal of computation by the students.

It was decided that the notions (i) - (iv) could appropriately form the subject of two laboratory sessions. In the first of these, called QUOTA SAMPLING, the students would "interview" a sample with prescribed composition; in the second, called STRATA SAMPLING, the students would experiment with various probability sampling designs.

The Population:

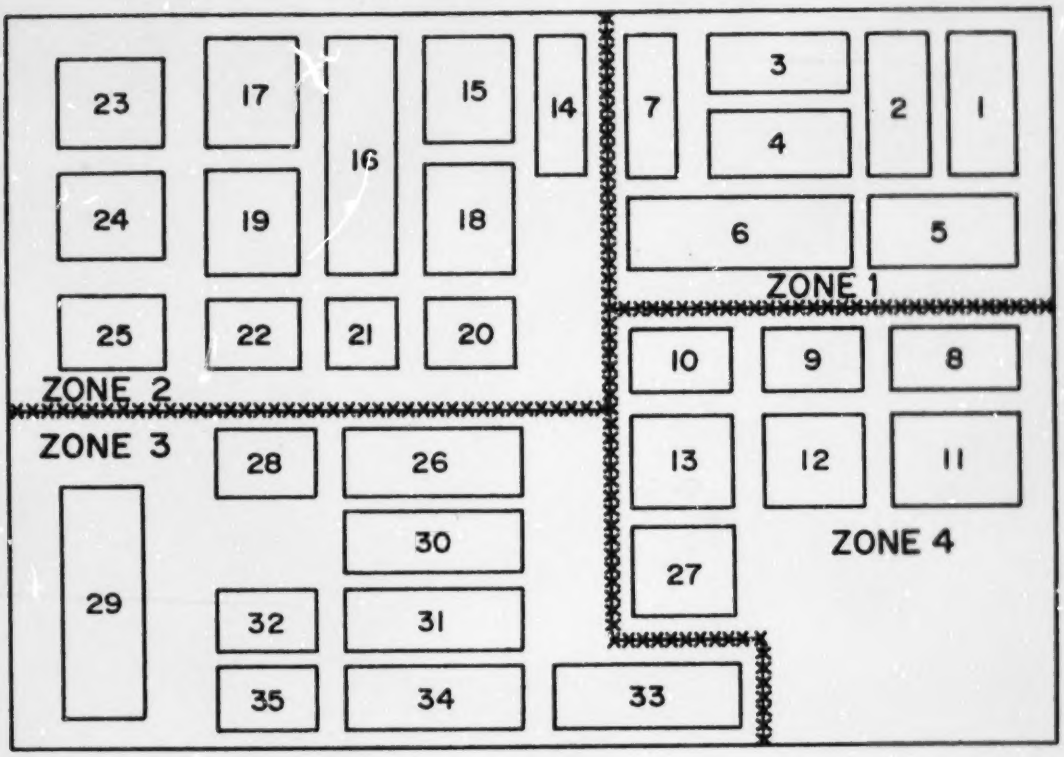
The first step in constructing the module was to set up a fictitious but relatively plausible population from which to sample. In the summer of 1975, Mr. Gerard C.F. Leung (at that time an undergraduate in Statistics) obtained tabulations of demographic data from the 1961 Census for four enumeration areas in the City of Waterloo. Using these (marginal) tabulations of sex, age group, occupation class, income, marital status, language and religious affiliation and origin, he created personal data for all inhabitants of 583 households, grouped into four zones (the enumeration areas). These data were entered in an SPSS file by the COMIT staff.

Following this, a map of the community (see Figure) was constructed with numbered street blocks, and each of the 583 households was assigned to an "address" in one of the blocks.

Each household was given a socioeconomic classification (W = high income, M = middle income, P = low income), and a block would usually contain households of one classification only.

The community included four apartment buildings with 24 units each, and four townhouse blocks with 12 units each. All other households corresponded to single family dwellings.

The population was set up so that, in its present version, occupation and income are related fairly believably. Also, family size and age of household head are associated fairly strongly with socioeconomic class (W,M or P)



BLOCK MAP OF FICTITIOUS COMMUNITY

and building type. That is, the P areas contain proportionately more elderly people and people with large families; the apartments are usually occupied by one or two adults, while the town houses tend to contain young families with children. Thus the location of a sample and its socioeconomic composition tend to affect its representativity as far as such characteristics as occupation and family size are concerned.

The Quota Sampling Exercise:

In QUOTA SAMPLING, each student is given a set of "quotas" or constraints on the composition of the sample he or she is to select. For example, a typical assignment for a sample of 15 people might be the following.

TABLE 1. A typical quota assignment

	Age Group		Marital Status		Social Class			Totals
	20-34	35-54	married	single	W	M	P	
Males	3	4	6	1	2	3	2	7
Females	3	5	6	2	2	4	2	8
								15

At the terminal the student is shown a schematic block map of the town, where each block is labelled with the number of dwelling units which are so far not sampled by the student, and the building types and socioeconomic classifications which are present in the block. The student points to a block with the light pen, and is shown a blow-up of the block with buildings indicated. Touching one of these will produce a "map" of individual units if the structure is an apartment building. Once an individual unit (apartment or house) is touched, the screen shows a list of the occupants and the demographic data for each, including sex, age group, marital status and occupation. The head of the household is designated as such. The student examines the list to see if there is any one whose characteristics fit the requirements of the sample; towards the end of the exercise, when 14 of the 15 people have been found, these requirements will be very specific, e.g. the student may need a W-class married male aged 35-54.

When a suitable 'interviewee' is found, the student points to the person, who may or may not be at home. (Which of these obtains is determined by a random number generated each time a selection is made.) If the person is home, he or she may refuse to answer questions or may be incompetent to answer questions. Otherwise, the call is considered successful. As the exercise is set up at present, the student is given an "interviewer's work sheet" on which to keep track of the sample composition and on which to record, for each successful call, the address and the interviewee's answers to three questions:

Q1: Would you vote for a woman for a high government office?

Q2: Do you believe that the government should do something to control the

amount of T.V. violence we see?

Q3: What is the occupation of the household head?

The answers to questions Q1, Q2 (see Table 2) are predetermined for every adult aged 20-54 in the population, and their distribution is dependent on own occupation except for married women, where the distribution is dependent on spouse's occupation. (Max Saltzman, the member of parliament has found that among his constituents in Waterloo, if husband and wife each fill out a mailed questionnaire on political issues the answers will very often be identical!)

Five students taking the course in the Winter of 1978 were each assigned a sample of size 15, in such a way that the total sample of 75 would be "representative" of the population aged 20-54 with respect to the quota "controls" (sex, age group, marital status and socio-economic classifications). Parts of the samples of two students were lost, so that the final sample size was less than 75. Table 2 shows the results of the quota sample. At least two noteworthy points emerged for discussion in class:

(i) Although the students were told to disperse their samples over the whole town, most of them found it difficult to do this and keep track of the sample composition at the same time; thus Zones 1 and 4, for example, were undersampled. Similarly, the students tended to avoid multiple family dwellings.

(ii) Zones 1 and 4 happened to be zones containing the largest proportions of "Yes" answers to Q2. The location bias in (i) undoubtedly contributed to the fact that the survey results estimate (wrongly) a preponderance of "No" answers.

(At the same time, with respect to (ii) it can be shown that a simple random sample of size 75 could easily yield the same 'incorrect' results. It is to be hoped that in future years classes will be large enough to make the results of the exercise possibly more conclusive.)

The students were graded according to the correctness of their sample compositions, the 'dispersion' of their samples, and their coding of responses to Q2, which were given in the form of sentences, like "Of course not, but

TABLE 2: Results from Quota Sampling Experiment

	Sample Totals						Sample Percentages	True Population Percentages
Student:	JY	PH	FY	MW	JD	All		
Occupation Class of Household Head								
Managerial	3	4	4	1	4	16	21.9	22.7
Professional	3	4	2	1	0	10	13.7	16.3
Clerical	1	1	0	3	2	7	9.6	10.9
Sales	0	2	1	1	2	6	8.2	8.2
Service	0	0	0	0	0	0	0	3.2
Craftsman	5	3	5	6	6	25	23.2	26.8
Labourer	0	0	0	0	0	0	0	4.0
Oth..	3	1	1	3	1	9	12.4	7.9
	15	15	13	15	15	73	100.0	100.0
Answer to Q1								
Yes	10	9	10	10	10	49	68.0	74.3
No	3	4	1	4	5	17	23.6	20.5
Don't Know	2	2	1	1	0	6	8.4	5.2
	15	15	12	15	15	72	100.0	100.0
Answer to Q2								
Yes, strongly	2	3	2	1	2	10	13.9	17.0
Yes	6	5	3	2	3	19	26.4	34.9
No opinion	2	0	2	5	3	12	16.7	11.2
No	4	6	4	5	6	25	34.7	26.0
No, strongly	1	1	1	2	1	6	8.3	10.9
	15	15	12	15	15	72	100.0	100.0
Location								
Zone 1	5	1	1	-	1	8	13.8	18.9
Zone 2	3	3	3	-	4	13	22.4	19.2
Zone 3	7	6	2	-	7	22	37.9	31.6
Zone 4	0	5	7	-	3	15	25.9	30.3
	15	15	13	-	15	58	100.0	100.0
Building Type								
Townhouse	0	0	0	-	0	0	0.0	8.2
Apartment	1	0	0	-	1	2	3.4	16.5
Other	14	15	13	-	14	56	96.6	75.3
	15	15	13		15	58	100.0	100.0

they ought to do something about all them dirty soap operas" (no,strongly). They found the exercise itself reasonably enjoyable and not difficult (according to one student, not difficult enough). They seemed to be quite interested in the results of the combined sample.

The Probability Sampling Exercise:

The exercise called STRATA SAMPLING is more technical in nature. Here the student is not to "interview" occupants of the households, but is required to "design" a probability sample for gathering household characteristics and to "test" the design.

The first thing that the student is asked to do is to "stratify" the blocks of the town into non-overlapping "strata". He or she may designate the town as one large stratum, or may stratify the town into the four zones, or may assign blocks to strata in any other way. The sample of households will then be the union of subsamples, one from each stratum. The student is asked to specify the number of households to be selected in each of the subsamples. He or she must then specify the sampling design to be used for the subsamples, and has four options (see [1] for definitions of the terminology):

A: simple random sampling of households

B: systematic sampling of households

C: two stage sampling, in which a simple random sample of blocks is selected first, and the households to be sampled are selected from these blocks

D: further stratification by building type, followed by simple random sampling within strata.

If C or D is chosen, there are more detailed sample size specifications to be made.

Once the design is completely specified, the program implements the design once, and displays estimates of quantities like the proportion of households with head having a specified occupation, or the average number of children aged 0-4 per household; together with these are shown estimated standard errors of estimation. It also prints out the true population value

for each quantity and the theoretical standard error of estimation under the design.

In the associated laboratory assignment in the Winter of 1978 the students were asked to try out the various options, with design parameters given to them, for a sample of total size 75. They were then asked to examine the results to see if they were in accordance with theory. For example, theoretically a dispersed sample (systematic or stratified by zone and allocated to the zones proportionally) should be relatively efficient, and tend to have lower theoretical standard errors than a clustered sample (two stage with several households chosen from each sampled block). The students were asked to compare the performance of these samples with that of the quota sample in estimating the distribution of occupation of household head. Finally, they were asked to design a sample which would be more "efficient" than any of the prescribed ones for estimating the average number of preschool children per household.

An unexpected feature of this exercise was the finding that as in real life, there were results for some of the prescribed designs which did not apparently agree with theory! Apart from some confusion, the effects of this were largely beneficial, since the author was forced to spend a good deal of time and effort trying to convince the students that the theory could be relied upon in general.

Conclusions:

Because of the small size of the class on which the module was tested, and because of difficulties encountered with the operation and scheduling of the second exercise, conclusions about the effectiveness of the module as a teaching tool must be tentative. Continued use of the module in future terms is planned, however, and more definite conclusions should be available at a later date.

Happily it appeared by the end of the second exercise that the notions (i), (iii) and (iv) mentioned in the Introduction had been conveyed to most of the students (more could have been done with (ii), in a classroom Monte Carlo study, if more time had been available). In addition, the impression gained was that this sort of simulation of survey situations could

be especially valuable as a device for enrichment of the curriculum, as well as a means of reinforcing concepts which have been encountered in lectures and books.

Assuming this to be the case, the author feels that the work described here, particularly the STRATA SAMPLING part, could be developed further. Obvious improvements would follow the direction of making the simulation more realistic; for example, formulae for measurement of sampling costs associated with the various options could be incorporated, and students could be asked to take travelling and interviewing costs into account when developing their designs. At the same time, as became very clear in the course of the module development, the larger the number of choices to be made available to student and instructor in the simulated survey situation, the more intricate and costly the job of writing and making available the associated software. Thus the author is grateful to the COMIT organizers and staff for making this experiment possible, and particularly to the programmer, Mr. G. Suggitt, who worked for many long hours to see the project through most capably.

Reference:

[1] S. Sudman, Applied Sampling, Academic Press, New York, 1976.